CLAIMS

We claim:

1. A system for the efficient control and capture of dynamic database content comprising:

a computer system having a storage means for facilitating the retention and recall of dynamic database content, said computer system having a communications means for performing bi-directional communications between said computer system and a network;

an executory module interfaced to said computer system, said executory module controlling said storage means of said computer system and said communications means of said computer system, said executory module directing said system to a plurality of pre-selected network sites;

a capture module being in communication with said executory module, said capture module facilitating selection of said plurality of network sites associated with a query submitted by a user of said system.

2. The system of claim 1, further comprising:

a query input means for receiving the query from a user and transferring said query to said capture module.

3. The system of claim 1, further comprising:

a query input means for receiving a plurality of queries from a user and transferring the plurality of queries to said capture module.

4. The system of claim 3, wherein said query input means

includes an input module selected from the group of input modules consisting of a keyboard operationally coupled to said computer system, a mouse operationally coupled to said computer system, a data input device capable of converting action of a user to a machine readable query, a data file transferred as an electrical signal to said computer system, a data file transferred as an optical signal to said computer system, a data file written to memory in said computer system accessible by said capture module, and a data file written to a storage media accessible by said capture module.

5. The system of claim 1, further comprising:

a database search listing for providing said capture module with a listing of a plurality of pre-selected databases to use for submission of a query.

6. The system of claim 5, wherein said database search listing further comprises at least one information field containing information for formatting queries submitted to each one of said plurality of pre-selected databases.

7. The system of claim 5, wherein said database search listing further comprises at least one information field conveying information for formatting of a query associated with at least one of said plurality of pre-selected databases.

8. The system of claim 1, wherein said executory module further comprises a network connectivity portion for establishing and maintaining bi-directional connectivity between said computer system and the network for facilitating the transmittal of at least one query to at least one site on said network.

9. The system of claim 1, wherein said executory module

further comprises a network connectivity portion for establishing and maintaining bi-directional connectivity between said computer system and a network for facilitating the transmittal of at least one query to at least one external network, said network connectivity portion also being for establishing and maintaining bi-directional connectivity between said computer system and an internal network to permit queries from a user to be directed databases on both internal and external networks.

10. The system for the efficient control and capture of dynamic database content of claim 9, wherein said network connectivity portion uses a plurality of sockets to establish and maintain bi-directional connectivity with the network.

11. A system for the efficient control and capture of dynamic database content comprising:

a computer system having a storage means for facilitating the retention and recall of dynamic database content, said computer system having a communications means for performing bi-directional communications between said computer system and a network;

an executory module interfaced to said computer system, said executory module controlling said storage means of said computer system and said communications means of said computer system, said executory module directing said system to a plurality of pre-selected network sites;

a capture module being in communication with said executory module, said capture module facilitating selection of said plurality of network sites associated with a query submitted by a user of said system;

a query input means for receiving a plurality of queries from a

user and transferring the plurality of queries to said capture
module;

a database search listing for providing said capture module a
listing of a plurality of pre-selected databases to use for submission
of a query;

said database search listing further comprising at least one
field conveying information for formatting of a query associated
with at least one of said plurality of pre-selected databases;

said executory module further comprising a network
connectivity portion for establishing and maintaining bi-directional
connectivity between said computer system and a network for
facilitating the transmittal of at least one query to at least one
external network, said network connectivity portion also being for
establishing and maintaining bi-directional connectivity between
said computer system and an internal network to permit queries
from a user to be directed databases on both internal and external
networks;

said network connectivity portion using a plurality of sockets
to establish and maintain bi-directional connectivity with the
network.

12. The system of claim 11, further comprising:
a query queue for storing queries from said query input means
until said queries are transferred to said executory module.

13 The system of claim 12 wherein said capture module
transfers each one of said queries to said executory module as a
thread.

14. The system of claim 12, wherein said capture module uses
a plurality of threads to transfer queries to said executory module

to thereby establish multiple coexisting sequential flows of control between said capture module and said executory module.

15. The system of claim 11, further comprising:

a query queue for storing queries from said query input means until each one of said queries are transferred to said executory module; and

a thread manager for the creation, management, and termination of a plurality of threads between said capture module and said executory module, each one of said plurality of threads transmitting a query from said capture module to said executory module and transmitting a reply from said executory module to said capture module received through one of said sockets from one of said databases.

16. The system of claim 15, wherein said thread manager terminates one of said sockets when one of said databases has completed responses to a series of queries to said database such that each thread terminates upon completion of queries and responses associated with a single database.

17. The system of claim 15, wherein said thread manager further comprises:

a simultaneous thread count parameter received from said capture module, said simultaneous thread count parameter being used by said thread manager to set an upper bound for a number of simultaneously coexisting threads forming a plurality of threads;

a thread creation and termination portion adapted for interacting with an operating system for generating a thread associated with one of said databases to be queried;

a plurality of monitoring portions, each monitoring portion

being associated with one of said plurality of threads, said monitoring portion determining a termination point when all responses associated with each one of said plurality of queries directed to an associated one of said plurality of databases have been returned;

said thread creation and termination portion terminating one of said plurality of threads when said termination point is reached by said monitoring portion associated with said one of said plurality of databases;

said thread creation and termination portion being operationally linked to said database search listing, said thread creation and termination module generating a new thread if additional databases are to be queried and said simultaneous thread count parameter has not been reached.

18. The system of claim 17, wherein said thread manager further comprises a timeout portion for determining a termination point for use by said thread creation and termination module when a database being queried fails to respond within a predetermined period of time.

19. The system of claim 11, wherein said executory module further comprises:

a document storage and retrieval portion for retaining documents returned in reply to an associated one of said plurality of queries submitted to each associated one of said plurality of databases.

20. The system of claim 19, wherein said document storage and retrieval portion further comprises:

a document storage module for retaining each one of a

plurality of documents as part of an indexed array for facilitating rapid retrieval of a document by the user.

21. The system of claim 20, wherein said document storage module stores each one said plurality of documents as part of a binary string of data, said plurality of documents being stored in a single file, each one of said plurality of documents being separately accessible.

22. The system of claim 19, wherein said document storage and retrieval portion further comprises:

an index portion for facilitating rapid recall of any one of said plurality of documents to be recalled from a document storage module.

23. The system of claim 22, wherein said index portion further comprises:

a B-tree as an indexing structure for said plurality of documents;

a plurality of core version uniform resource locators, each core version URL providing a path back to a source document from an associated one of the databases queried allowing reconnection to the database which provided the source document.

24. The system of claim 19, wherein said document storage and retrieval portion further comprises:

a uniform resource locator module for retaining uniform resource locator associated with each one of said plurality of documents returned by each one of said plurality of databases.

25. The system of claim 19, wherein said document storage and retrieval portion further comprises:

an entity tag portion for retaining an entity tag associated with each one of said plurality of documents returned by each one of said plurality of databases.

26. The system of claim 19, wherein said document storage and retrieval portion further comprises:

a record related information portion containing parametric information associated with each one of said plurality of documents, said record related information portion being for facilitating analysis of each one of said plurality of documents.

27. The system of claim 26, wherein said record related information portion further comprises:

a plurality of offset segments, each one of said plurality of offset segments being a representation of a starting point for an associated one of said plurality of documents in a document storage module;

a plurality of length segments, each one of said plurality of offset segments being a representation of a length for an associated one of said plurality of documents in said document storage module;

a plurality of last-time-checked segments, each one of said plurality of last-time-checked segments being a representation of the last known occurrence of collecting an associated one of said plurality of documents in said document storage module;

a plurality of hit segments, each one of said hit segments being a representation of a number of previous requests received for an associated one of said plurality of documents;

a plurality of highest-score segments, each one of said highest-score segments being a representation of best results obtained for an associated one of said plurality of documents in

said document storage module compiled through use of an arithmetic scoring operation;

a plurality of database segments, each one of said database segments being a representation of a search engine used to locate an associated one of said plurality of documents in said document storage module.

28. The system of claim 27 wherein each one of said plurality of offset segments, length segments, last-time-checked segments, hits segments, highest score segments, and database segments comprises a 32 bit representation, each one of said segments being stored as part of an array, each one of said offset segments having a one to one correspondence with an index portion linking a stored version of each one of said plurality of documents with an associated one of said parametric information.

29. The system of claim 27 wherein each one of said plurality of offset segments, length segments, last-time-checked segments, hits segments, highest score segments, and database segments comprises a 64 bit representation, each one of said segments being stored as part of an array, each one of said offset segments having a one to one correspondence with an index portion linking a stored version of each one of said plurality of documents with an associated one of said parametric information.

30. The system of claim 19 wherein said document storage and retrieval portion further comprises a version control portion for recording a version identification for said document storage and retrieval portion, said version control portion allowing the user to verify configuration attributes of said document storage and retrieval portion.

31. The system of claim 19, wherein said document storage and retrieval portion further comprises:

a document storage module for retaining each one of a plurality of documents as part of an indexed array for facilitating rapid retrieval of a document by the user;

wherein said document storage module stores each one said plurality of documents as part of a binary string of data, said plurality of documents being stored in a single file, each one of said plurality of documents being separately accessible; and

wherein said document storage and retrieval portion further comprises an index portion for facilitating rapid recall of any one of said plurality of documents to be recalled from a document storage module;

wherein said index portion further comprises:

a B-tree as an indexing structure for said plurality of documents;

a plurality of core version uniform resource locators, each core version uniform resource locator providing a path back to a source document from an associated one of the databases queried for allowing recommendation to the database which provided the source document;

wherein said document storage and retrieval portion further comprises:

a uniform resource locator module for retaining an associated uniform resource locator associated with each one of said plurality of documents returned by each one of said plurality of databases;

an entity tag portion for retaining an entity tag associated with each one of said plurality of documents returned by each one of said plurality of databases;

a record related information portion containing parametric information associated with each one of said plurality of documents, said record related information portion being for facilitating analysis of each one of said plurality of documents;

a plurality of offset segments, each one of said plurality of offset segments being a representation of a starting point for an associated one of said plurality of documents in a document storage module;

a plurality of length segments, each one of said plurality of offset segments being a representation of a length for an associated one of said plurality of documents in said document storage module;

a plurality of last-time-checked segments, each one of said plurality of last-time-checked segments being a representation of the last known occurrence of collecting an associated one of said plurality of documents in said document storage module;

a plurality of hit segments, each one of said hit segments being a representation of a number of previous requests received for an associated one of said plurality of documents;

a plurality of highest-score segments, each one of said highest-score segments being a representation of best results obtained for an associated one of said plurality of documents in said document storage module compiled through use of an arithmetic scoring operation;

a plurality of database segments, each one of said database segments being a representation of a search engine used to locate an associated one of said plurality of

30

documents in said document storage module; and

a version control portion for recording a version

identification for said document storage and retrieval portion,

said version control portion allowing the user to verify

configuration attributes of said document storage and

retrieval portion.

32. The system of claim 31, wherein each one of said plurality
of offset segments, length segments, last-time-checked segments,
hits segments, highest score segments, and database segments
comprises a 32 bit representation, each one of said segments being
stored as part of an array, each one of said offset segments having a
one to one correspondence with an index portion linking a stored
version of each one of said plurality of documents with an
associated one of said parametric information.

33. The system of claim 31, wherein each one of said plurality
of offset segments, length segments, last-time-checked segments,
hits segments, highest score segments, and database segments
comprises a 64 bit representation, each one of said segments being
stored as part of an array, each one of said offset segments having a
one to one correspondence with an index portion linking a stored
version of each one of said plurality of documents with an
associated one of said parametric information.

34. The system of claim 31, wherein said capture module uses
a plurality of threads to transfer queries to said executory module
to thereby establish multiple coexisting sequential flows of control
between said capture module and said executory module;

a query queue for storing queries from said query input means
until each one of said queries are transferred to said executory

module;

a thread manager for the creation, management, and termination of a plurality of threads between said capture module and said executory module, each one of said plurality of threads transmitting a query from said capture module to said executory module and transmitting a reply from said executory module to said capture module received through one of said sockets from one of said databases;

wherein said thread manager terminates one of said sockets when one of said databases has completed responses to a series of queries to said database such that each thread terminates upon completion of queries and responses associated with a single database;

wherein said thread manager further comprises:

a simultaneous thread count parameter received from said

executory module, said simultaneous thread count parameter being used by said thread manager to set an upper bound for a number of simultaneously coexisting threads forming a plurality of threads;

a thread creation and termination portion adapted for interacting with an operating system for generating a thread associated with an associated one of said databases to be queried;

a plurality of monitoring portions, each monitoring portion being associated with one of said plurality of threads, said monitoring portion determining a termination point when all responses associated with each one of said plurality of queries directed to an associated one of said plurality of databases have been returned;

32

said thread creation and termination portion terminating
one of said plurality of threads when said termination point is
reached by said monitoring portion associated with said one
of said plurality of databases;

wherein said thread creation and termination portion is
operationally linked to said database search listing, said thread
creation and termination module generating a new thread if
additional databases are to be queried and said simultaneous thread
count parameter has not been reached; and

wherein said thread manager further comprises a timeout
portion for determining a termination point for use by said thread
creation and termination module when a database being queried
fails to respond within a predetermined period of time.

35. The system of claim 34, further comprising:

a redirected URL handler portion for following redirection of
an URL through a plurality of redirections to an ultimate
destination without maintaining intermediate pages, said redirected
URL handler portion providing an URL for the user of the ultimate
document.

36. The system of claim 31, further comprising:

a redirected URL handler portion for following redirection of
an URL through a plurality of redirections to an ultimate
destination without maintaining intermediate pages, said redirected
URL handler portion providing an URL for the user of the ultimate
document.

37. The system of claim 36, wherein said document storage
and retrieval portion further comprises:

a source URL portion for retaining a URL associated with of

33

each one of a plurality of databases providing documents to said system for the efficient control and capture of dynamic database content;

an ultimate URL portion for retaining a URL associated with each one of a plurality of ultimate destinations obtained through said redirected URL handler portion.

38. The system of claim 11, further comprising:

a redirected URL handler portion for following redirection of an URL through a plurality of redirections to an ultimate destination without maintaining intermediate pages, said redirected url handler portion providing an URL for the user of the ultimate document.

39. The system of claim 34, further comprising:

a document aging portion for determining if a current version of a document is available from said document storage and retrieval portion or if the document must be retrieved from another source through said network connectivity portion.

40. The system of claim 39, wherein said document aging portion further comprises:

an aging parameter for selecting a predetermined maximum age for a document to be considered current;

an age module for determining when the document was retrieved from a source and if said age parameter has been exceeded;

a modification module adapted for interrogating a server about any changes made to the document since the document was previously retrieved.

41. The system of claim 40, wherein said modification module

uses an entity tag to determine if the document has been modified, said modification module also using a last-modified-since tag if the server does not support use of the entity tag.

42. The system of claim 19, further comprising:

a document aging portion for determining if a current version of a document is available from said document storage and retrieval portion or if the document must be retrieved from another source through said network connectivity portion;

wherein said document aging portion further comprises:

an aging parameter for selecting a predetermined maximum age for a document to be considered current;

an age module for determining when the document was retrieved from a source and if said age parameter has been exceeded;

a modification module adapted for interrogating a server about any changes made to the document since the document was previously retrieved;

wherein said modification module uses an entity tag to determine if the document has been modified, said modification module also using a last-modified-since tag if the server does not support use of the entity tag.

43. The system of claim 42, wherein said executory module further comprises:

a scoring portion for evaluating each one of said plurality of documents for relevance against a query provided by the user, said scoring portion providing a numeric representation of relevance for the user.

44. The system of claim 31, wherein said executory module

further comprises:

a scoring portion for evaluating each one of said plurality of documents for relevance against a query provided by the user, said scoring portion providing a numeric representation of relevance for the user.

45. The system of claim 11, wherein said executory module further comprises:

a scoring portion for evaluating each one of said plurality of documents for relevance against a query provided by the user, said scoring portion providing a numeric representation of relevance for the user.

46. A method for efficiently controlling and capturing dynamic database content comprising:

obtaining a query from a user, said query to be used to search for relevant information in a specified area;

pre-processing said query;

routing said query to a plurality of said databases, each one of said databases using said query to locate relevant information for the user;

obtaining results from each one of said plurality of databases;

processing said results;

storing said results for use and analysis by the user.

47. The method of claim 46, wherein said step of pre-processing further comprises:

obtaining a database search listing, said database search listing containing a plurality of databases determined to be relevant for specified domain areas;

comparing said query to said database search listing to

determine the best matches from all candidate databases;

selecting the most relevant databases to submit said query from the user.

48. The method of claim 46, wherein said step of pre-processing further comprises:

invoking a thread manager to facilitate the submission of said query to said plurality of databases, said thread manager coordinating communication between an executory module and a capture module, said executory module controlling bi-directional communications with said databases;

obtaining a thread count parameter for limiting the maximum number of coexisting threads allowed by said thread manager;

establishing a plurality of threads between said capture module and said executory module such than a plurality of coexisting sequential control flows are created between said capture module and said executory module;

establishing a plurality of sockets for said executory module to generate bi-directional connectivity between said executory module and said plurality of appropriate databases;

assigning each one of said plurality of databases to a specified socket;

obtaining a query from a queue;

routing said query from said capture module through an associated one of said plurality of threads to said executory module and from said executory module through an associated one of said plurality of sockets to an associated one of said plurality of appropriate databases;

waiting for a response from said one of said plurality of appropriate databases;

capturing results returned by each one of said plurality of

databases for processing and storage.

49. The method of claim 46, wherein said step of pre-processing further comprises:

building a query queue to store each one of a plurality of queries until each one of said queries is routed to said plurality of databases;

queuing said plurality of queries to be submitted to each one of said plurality of databases;

routing each query in turn to each one of said plurality of databases

50. The method of claim 46, wherein said step of obtaining results further comprises:

capturing links to relevant documents returned from one of said databases via a results page, said links providing a URL path to obtain each one of said documents;

examining each one of said URLs to determine if a redirection is used to reach an ultimate location for any one of said documents;

following each one of said URLs to determine if subsequent redirects are used;

capturing one of said documents from said database;

capturing an ultimate URL portion associated with said document;

capturing a source URL portion identifying which one of said databases was a referring source of said document;

deleting any intervening URLs between said source url portion and said ultimate URL portion.

51. The method of claim 46, wherein said step of obtaining results further comprises:

obtaining a aging parameter from the user, said aging parameter defining the maximum age of a document to be considered current;

checking each URL for a document returned as a result by one of said databases with documents stored in a document storage and retrieval portion, said document storage and retrieval portion storing all documents captured during a search;

determining if said document is located in said document storage and retrieval portion;

downloading said document if said document is not located in said document storage and retrieval portion;

examining record related information for said document if said document is located within said document storage and retrieval portion;

verifying an age of said document from said record related information;

comparing said age of said document with said aging parameter;

querying a server associated with said document for the existence of modifications to said document if said document is older than the maximum age determined by said aging parameter;

downloading said document if said document has been modified since the last time the document was downloaded.

52. The method of claim 46, wherein said step of processing further comprises:

obtaining an exclusion list detailing terms, phrases, and sources to be excluded from responses to said query;

comparing each responses associated with said query to said exclusion list;

deleting any responses which match terms, phrases, and

39

sources on said exclusion list.

53. The method of claim 46, wherein said step of processing further comprises:

obtaining an inclusion list detailing terms, phrases, and sources which must be contained in each response to said query;

comparing each response associated with said query to said inclusion list;

deleting any response which does not include terms, phrases, and sources on said inclusion list.

54. The method of claim 46, wherein said step of processing further comprises:

scoring each document returned from each of said databases for relevance compared to said query;

recording a plurality of scores, each score associated with one of said responses from each one of said plurality of databases;

obtaining a minimum acceptable score for narrowing the documents to be considered;

comparing each one of said plurality of scores with said minimum acceptable score;

deleting documents which have an associated score numerically less than said minimum acceptable score.